경영학석사학위논문

User Behavior Analysis(UBA) of a MMORPG Game Based on the Flow Theory: Reflecting on Factors Learned from a Two-Stage Least Squares Regression Model

몰입이론(Flow Theory)에 따른 온라인 게임 내 사용자 행태변화 분석

2017 년 2 월

서울대학교 대학원

경영학과 경영학 전공

김 은 찬

User Behavior Analysis (UBA) of a MMORPG Game Based on the Flow Theory: Reflecting on Factors Learned from a Two-Stage Least Squares Regression Model

지도교수 유 병 준

이 논문을 경영학 석사학위논문으로 제출함

2017년 1월

서울대학교 대학원

경영학과 경영학 전공

김은찬

김은찬의 석사학위논문을 인준함

2017 년 1월

위 원 장 장 정 주 (인)

부 위 원 장 노 상 규 (인)

위 원 유 병 준 (인)

Abstract

User Behavior Analysis(UBA) of a MMORPG Game Based on the Flow Theory: Reflecting on Factors Learned from a Two-Stage Least Squares Regression Model

Eunchan Kim

College of Business Administration

The Graduate School

Seoul National University

In this study, by analyzing cross-sectional online game data, it is aimed to examine how users get influenced on their behaviors and how each factor affects each other based on the Flow Theory. In order to understand this purpose, basic information and recent studies on the Flow theory are fully covered. Next, few interesting analyses on users' behaviors with selected variables are studied reflecting on factors learned from a two-stage least squares regression model. Since

a number of studies related to the Flow Theory were mainly done with primary data, finally, research achievements of this study is included with its controversial limitations and further research ideas.

Keywords: MMORPG, Online Games, Two-Stage Least Squares, Flow Theory, User Behavior Analysis

Student ID Number: 2015 - 20594

Table of Contents

CHAPTER 1. Introduction

It is fairly said that a new generation of RPG (Role Playing Game) has begun because of MMORPG (Massively Multiplayer Online Role Playing Game). It is not only because a distribution rate of an internet made its highest record recently but also because its systems develop rapidly (Yoo, 2013). Some people even say that this kind of thriving online game environment will continue catching up with trend. In this sense, online game companies figure out many ways to survive through severe competitions. They offer a large number of promotions for game users, use aggressive marketing strategies, and hold active events in both offline and online circumstances.

However, above all things the most important for online game companies is to apprehend their users' attributes, preferences, and behaviors. In short, knowing their users deeply is getting more important than ever. By doing so, online game companies could balance game difficulties out based on the analyses of game users. Furthermore, these analyses could give another chance to improve companies' business circumstances and those analyses are full of suggestions. The business implications of users' behavioral analyses are numerous. In the example at

the top, online game companies could gain more profits from their analyses on users.

In this study, by analyzing cross-sectional online game data, it is aimed to examine how users get influenced on their behaviors and how each factor affects each other based on the Flow Theory. In order to understand this purpose, basic information and recent studies on the Flow theory are fully covered. Next, few interesting analyses on users' behaviors with selected variables are studied reflecting on factors learned from a two-stage least squares regression model. Since a number of studies related to the Flow Theory were mainly done with primary data, finally, research achievements of this study is included with its controversial limitations and further research ideas.

CHAPTER 2. Theoretical Background

The Flow Theory is first academically formed in 1970 by Csikszentmihalyi M., a Hungarian psychologist. The author introduced a concept of the flow theory for the first time to account for how and why people concentrate on doing something and get motivated. According to the article published in 1978, it gives reasonable explanations about concentrations and self-consciousness in a private individual. A concentrated person feels happiness and greatness in what he or she actually doing, and it is optimal experience of a human being.

In other words, concentrating on something is maximum level of a psychological experience as a personal action. The author also mentioned that this emotional and psychological state of human can be found with one's satisfaction feeling, and it is often formed in a situation between a challengeable and acceptable stage (Csikszentmihalyi, 1978).

There have been many studies based on the Flow Theory since the theory was established. Lots of studies mostly focused on one's addiction on doing something. Speaking of which, it is very controversial to find out whether a person is concentrated or addicted. Therefore, how a human being

gets concentration or addiction has always been the subject of debate by various scholars. Because it is hard to separate concentration and addiction, some scholars say it ought to be a very subjective analysis and interpretation. On the other hand, the other scholars say that a personal retention of feeling is considerably clear so that a condition of one's immersion and addiction is easily separated.

In spite of these controversies by numerous scholars, there have been lots of advances in the academia. Some studies were following the classical point of view on the theory. Some studies got out of the existing standpoint to expand the study area from the real world to the person's online status. During early 2000's, reflecting the sudden changing internet era, there were various studies on internet addiction and abuse (Chou et al., 2000; Morahan-Martin et al., 2000). As the internet environment has been introduced, people often got addicted; therefore, studies on internet addiction problems were important enough to arouse people's attention to the internet abuse at that time.

After the internet transition period, studies on users of online game have been popular. Among various online game studies, there are a few studies which made their remarkable development on the academia (Ting et al., 2003; Hsu et al., 2004). These scholars introduced the methodologies to

analyze behavioral changes based on the Flow Theory for the first time. Hsu, the author, even extended the TAM (Technology Acceptance Model) into the online game status in order to explain and prove why people keep playing online games. It provided a background for the intelligent study of online game users' behavioral changes.

In late 2000's, Thanks to the extended study area of the Flow Theory, scholars even tried to expand study perspectives into the online game status (Holt et al., 2009; Wen et al., 2009). Holt and his colleague analyzed users' behavioral changes with primary survey data of online multi-playing game. The author studied the behavioral changes of differences between adults and adolescences. Therefore, it was evaluated as very distinguishing study then.

Another online game study was a prediction study on MMORPG addiction (Wen et al., 2009). Wen and his colleagues tried to predict when a game user would get addicted. For a prediction, they needed to set an addiction level of playing game. It was very controversial because one's addiction is pretty hard to measure, or it is not measurable. Since they also used a survey methodology like other scholars, the study had its limitations for that.

As the interim history of the Flow Theory was covered with a few literature reviews, the Flow Theory has developed and expanded by a number of scholars. However, notwithstanding these efforts, there are numerous limitations of existing studies. First, the studies too often focused on addiction matters. Although the procedure of analysis was very detailed and well managed to show users' behavioral changes, existing studies sometimes failed to prove addiction or even lost their study directions. Furthermore, a personal addiction is still controversial to say whether he or she is addicted or not, there are so many reasons to avoid those kind of studies.

Second, almost all studies on the Flow Theory are based on survey data. It occasionally failed to be objective. Actually, several researches were proved to be false, and some were criticized by other scholars. Last but not least, the methodologies used in existing studies are too old to use and not diverse. Most studies used TAM (Technology Acceptance Model), Hierarchical Analysis, and clustering analysis. This matter results from the fact that existing studies have used primary data.

CHAPTER 3. Data Description and Hypotheses

- Basic Information of Data

The data which are analyzed in this study are from the MMORPG of a Korean game company. Since it is a closed data and contains very personal information, specific names of the game and its company are not provided in this study. Therefore, if anyone have a curiosity about data, please contact the author individually. The data are cross-sectional; therefore, it is presumed that the sampled data are representative sample of the whole standard data set.

- Data reprocessing procedure

First, because the given data set from the game company contains various forms including structured, semi-structured, and unstructured data, there needs to be a separation operation. Therefore, the text data are removed because this study mainly focuses on numeric character data. Text data are deposited for another research. Second, by looking over and plotting the numerical data, it is possible to separate linear and non-linear data. For easier analysis purpose, putting the natural logarithm or the square root of non-linear data. Third, in an attempt to select and deselect variable for the research, some of recent studies on the Flow Theory are referred

(Choi et al., 2004; Criffiths et al., 2004; Hsu et al., 2004; Nakamura et al., 2009).

- Selected Variables

Selected Variable	Content	Mean	Median	Min	Max
UserID	User Identification	-	-	-	-
Age	User Age	23.87	18.00	1.00	97.00
Gender	User Gender	1.34	1.00	2.00	1.00
TotalTime	Total Playing Time (sec.)	1.64e+08	2.83e+05	4.24e+11	1.00
BattleTime	Total Battle Engaged Time	13015	1125	0.00	2296875
RestTime	Total Resting Time	53964	7150	1.00	8657258
PartyCount	Party Connection Count	68.41	0.00	0.00	19994.00
PartyMember	Party Member Count	206.1	0.00	0.00	78999.00
Party	Whether in a Party or Not	0.487	0.00	0.00	1.00
PurchaseCount	Purchase Count	0.811	0.00	0.00	622.00
Purchase	Whether Purchase or Not	0.079	0.00	0.00	1.00

Table 1. Selected Variables

Each and every variable has been chosen through the basic flow theory. Each description of the chosen variable is as follows.

1) UserID is each individual's identification to differentiate one with another user.
2) Age is each user's age varied from 1.00 to 97.00.
3) Gender is each user's sex simply man (1.00) and woman (2.00).
4) TotalTime is time variable; the unit is second basis, and each value shows accumulated time of user's total playing.
5) BattleTime is also second basis unit, and it shows collected time of user's hunting and battling.

6) RestTime is second basis and it shows accumulated time of user's resting and chatting.
7) PartyCount is partying count value of how many party and connecting action the user joined.
8) PartyMember is counting value of how many party members the user played together when the user partying.
9) Party is binomial value showing whether user has a party connection experience or not.
10) PurchaseCount is count value of how many times the user actually purchases cash items.
11) Purchase is binomial value showing whether the user has a purchase experience or not.

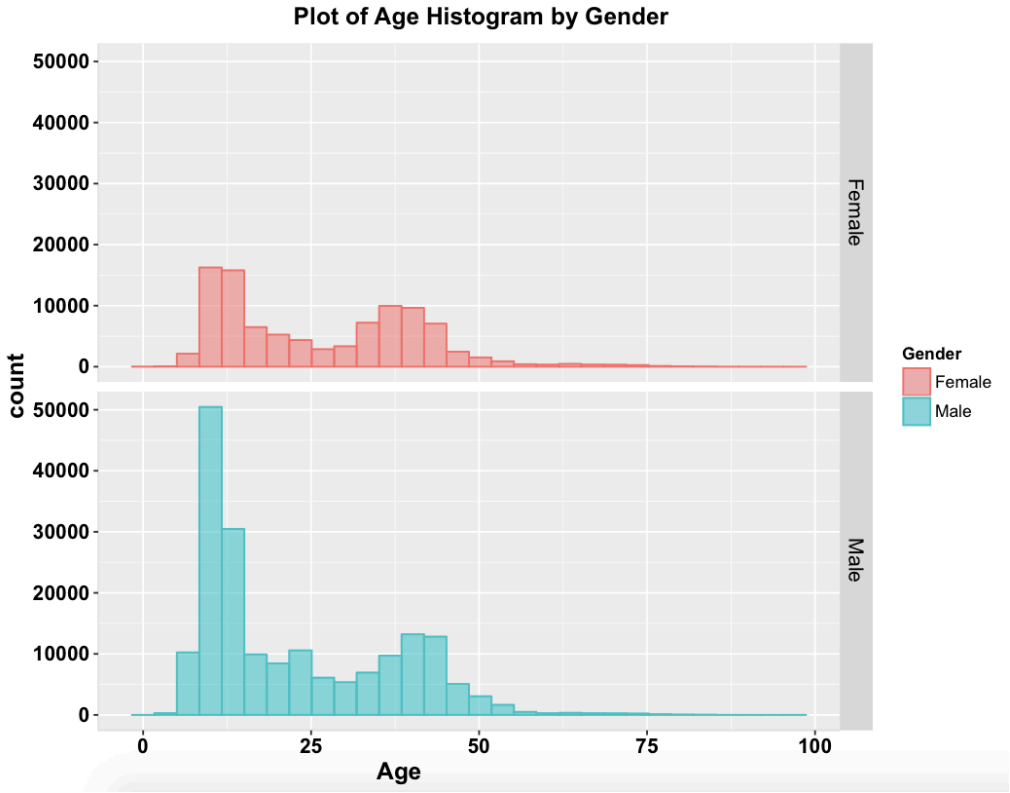

Figure 1. Age Histogram by Gender

The figure shows that the game players' distribution of population by gender. As it is shown, it can be depicted that comparatively young aged players are active players; on the other hand, old aged players are not. From this idea, the research focuses on the differences about why young and adult groups are showing unlike playing aspects. Therefore, it is set and planned to check each group and its attributes in following contents.

- Hypotheses

1. H1: Whole observation are most affected to keep playing the game by the purchase count variable among selected variables.

2. H2-1: The adult group gets more affected by social activities and networking (such as, having a party in this study) than the adolescence group; therefore, the adult group has high probability of an immersion symptom which causes continuous playing.
3. H2-2: The adolescence group gets more affected by purchasing activities than the adult group, which attracted the adolescences to keep playing the game.
4. H3: Social and networking activities would affect more purchasing probabilities than total playing time would affect those.

CHAPTER 4. Methods

The main purpose of this study is to find out what kinds of factors could feasibly affect an individual user's playing time in terms of one's concentration and dedication following by the Flow Theory. Therefore, total playing time of each user is set to be analyzed as a dependent variable; age, gender, purchase counts, and party counts are set to be independent variables. However, it is found that there is a possible endogenous issue between user's total playing time variable and user's purchase count variable. Therefore, in order to remove the endogeneity problem and minimize an impact of the problem, a two-stage least square model is adopted (Greene, 2008; Gujarati, 2009; Wooldridge, 2013).

From the first stage, the purchase count variable, "PurchaseCount" is set as a dependent variable; the user's total playing time variable, "TotalTime" is set as an independent variable in order to recalculate new "PurchaseCount" variable. Then, the recalculated purchase count variable is now usable for an analysis in the second stage. In the second stage, there are three main analyses and one additional analysis. In these analyses, it is very credible because those use the recalculated variable, which resolves possible endogenous issues.

During the second stage analyses, Multi-linear regression method and logistic regression method are used to analyze. The analyses discover not only what factors influence user's playing time but also what factors have an effect on user's purchases. In the procedure of Multi-linear regression, it is aimed to check the factors which affects playing time, and total observations are first analyzed to show how overall users behave. Then, it is set to analyze adult and adolescence groups so as to reveal the differences between those two groups. Last, in the logistic regression analysis, a user's purchasing experience binomial variable, "Purchase" is used for an additional analysis so as to find out effects of each factor such as, age, gender, total time, and party count.

CHAPTER 5. Models

1. Two-Stage Least Square Regression Analysis (Stage 1)
    a. $\mathrm{Y(PurchaseCount)} = \alpha + \beta 1 * X1(TotalTime) + \varepsilon$

As it is explained, in order to reduce an endogenous issue, the two-stage least square regression model is adopted. In the first stage of it, the purchase count variable, "PurchaseCount" is set as dependent variable; the total playing time of each user, "TotalTime" variable is set as independent variable at this time. It has a purpose to re-calculate a new(hat) "PurchaseCount" variable, and this new variable is gained through the process of the first stage.

2. Multiple Regression Analysis with the New Variable (Stage 2)

- Total: 284,194

    a. $\mathrm{Y(TotalTime)} = \alpha + \beta 1 * X1(Age) + \beta 2 * X2(Gender) + \beta 3 * X3(PurchaseCount) + \beta 4 * X4(PartyCount) + \varepsilon$

After gaining this new "PurchaseCount" variable from the first stage, the newly calculated purchase count variable is put for one of independent variables as it is supposed to be. In this model, it is supposed to

look over each independent variable's effect on the dependent variable with overall observarions. By doing it so, it is expected to what factors have an impact on the user's total playing time variable; which tells how and how much each and every user is focused on playing the game through the Flow Theory.

3. Multiple Regression Analysis with the New Variable (Stage 2) – Adult: 137,127
    a. $\mathrm{Y(TotalTime)} = \alpha + \beta 1 * X1(Age) + \beta 2 * X2(Gender) + \beta 3 * X3(PurchaseCount) + \beta 4 * X4(PartyCount) + \varepsilon$

While it is shown that the overall effects and relations among variables in the previous model, at this time it is needed to look over the hierarchical cluster for adult and adolescence groups. In the third model, an analysis of the adult group is studied.

4. Multiple Regression Analysis with the New Variable (Stage 2) – Adolescence: 147,067
    a. $\mathrm{Y(TotalTime)} = \alpha + \beta 1 * X1(Age) + \beta 2 * X2(Gender) + \beta 3 * X3(PurchaseCount) + \beta 4 *$

$$X4(PartyCount) + \varepsilon$$

In the fourth model, an analysis of the adolescence group is studied to have a closer look on differences of characteristics between adult and adolescence groups.

5. Additional Regression Analysis (Logistic Regression Model)

   a. $\ln(\mathrm{Y})(\mathrm{Purchase}) = \alpha + \beta 1 * X1(Age) + \beta 2 * X2(Gender) + \beta 3 * X3(TotalTime) + \beta 4 * X4(PartyCount) + \varepsilon$

In the last analysis model, an extra analysis is processed to show purchase probabilities in accordance with each variable. Therefore, the binomial variable, "Purchase" is set as a dependent variable. "Age", "Gender", "TotalTime", and "PartyCount" variables are set to be independent variables.

CHAPTER 6. Results of Analyses

1. Two-Stage Least Square Regression Analysis (Stage 1)

Dependent Variable: PurchaseCount		
Independent Variable	**Coefficient**	**Std. Error**
TotalTime	**1.042e-09*** **	4.851e-12
***p<0.001, ** p<0.01, * p<0.05, . p<0.1 F: 4.61e+04, R^2: 0.1396, AR^2: 0.1396		

Table 2.

The table is the result of a two-stage least square regression analysis model's the first stage. As it is mentioned earlier, because there needs a re-calculating process for the new purchase count variable, it can be checked that each variable is put inversely. Through this procedure, the newly calculated variable is achieved as the result.

2. Multiple Regression Analysis with the New Variable (Stage 2)

   - Total: 284,194

Dependent Variable: PlayTime		
Independent Variable	**Coefficient**	**Std. Error**
Age	**-52688**	226680
Gender	**-48411524*** **	6846104
PurchaseCount	**7247459*** **	567391
PartyCount	**3987370*** **	9246
***p<0.001, ** p<0.01, * p<0.05, . p<0.1 F: 6.558e+04, R^2: 0.48, AR^2: 0.48		

Table 3.

From the Table 3, it can be depicted that there are several factors affect playing time statistically significant. The age variable has negative effect on playing time variable, but the p-value is not significant enough. Gender variable shows that players whose gender is male are more into the game, and spend more time than female players spend. The purchase count variable seems that it affects playing time most among the selected independent variables; and it is statistically significant enough. The party count variable is secondly most affective variable on the playing time variable; and it is statistically strong enough to say.

3. Multiple Regression Analysis with the New Variable (Stage 2)

- Adult: 137,127

Dependent Variable: PlayTime		
Independent Variable	**Coefficient**	**Std. Error**
Age	**5536081*** **	619233
Gender	**-62742806*** **	12625717
PurchaseCount	**5544730*** **	817420
PartyCount	**4186163*** **	13345
***p<0.001, ** p<0.01, * p<0.05, . p<0.1 F: 3.432e+04, R^2: 0.50, AR^2: 0.50		

Table 4.

4. Multiple Regression Analysis with the New Variable (Stage 2)

- Adole: 147,067

Dependent Variable: PlayTime		
Independent Variable	**Coefficient**	**Std. Error**
Age	**2876890*** **	769730
Gender	**-19775968*** **	5024374
PurchaseCount	**19333466*** **	757430
PartyCount	**2826180*** **	12497
***p<0.001, ** p<0.01, * p<0.05, . p<0.1 F: 1.966e+04, R^2: 0.35, AR^2: 0.35		

Table 5.

From the Table 4 and Table 5, it is shown that there are two great differences between adolescence and adult groups. That is, when

individuals playing, the adult group gets more affected by party counts (interactions and networking in the game) than the adolescence group. However, the adolescence group gets more affected by purchase counts than the adolescence group when they are playing the game.

5. Additional Regression Analysis (Logistic Regression Model)

Dependent Variable: Purchase (Binary)		
Independent Variable	**Coefficient**	**Std. Error**
Age	**1.010e-03*** **	3.382e-05
Gender	**1.048e-03**	1.022e-03
TotalTime	**2.077e-11*** **	2.024e-13
PartyCount	**1.257e-01*** **	9.64e-04
***p<0.001, ** p<0.01, * p<0.05, . p<0.1 F: 1.966e+04, R^2: 0.35, AR^2: 0.35		

Table 6.

From the result of the additional analysis, there are several things to mention. First, the player's age affects the purchasing experience positively with significance of it. Second, the total playing time variable has effect on purchasing experience positively with very little coefficient value. Third, the party count variable affects purchase experience positively with a strong statistical significance. Last but not least, under control of other independent variable, the party count

variable affects the players' purchase experiences most; it actually shows that if a player participates in a party, the player has about 12.5% higher chances to purchase an item in the game.

CHAPTER 7. Conclusion and Discussion

- Conclusion

From the results of analyses, the study could bring out many interesting points, and there are several great outcomes to mention. First, as the study uses microscopic secondary data not using primary survey data, it improves objectivity and justification of the study itself. Which, of course, could be evaluated that this study actually expands viewpoints of the Flow Theory and avoids controversial scholars' subjective interpretation problems. Second, by using the two-stage least square regression methodology, it completely removes the endogenous issue. Since there could be a possible endogeneity problem between purchase counts and total playing time variables, that problem is captured in the beginning of the research, and the problem is wisely solved.

Third, although the study could be just comparative study of comparing how each independent variable affects dependent variable like the other researches have done, this study uses hierarchical comparing analysis by adult and adolescence groups. Therefore, it brings very interesting point that the adult group's total playing time is more affected by the party counts, while the adolescence group's total playing time is getting

more effects from the purchase counts. It seems that the adult group is more social in the game, while the adolescence group is not.

Last but not least, through the additional analysis of logistic model, the study seeks out a great result for a business implication. It is about the networking effect in the MMORPG; there needs more elements for networking components. Since one more party count of a user gets more than 12% higher purchasing probability, networking and interactions with another user in the MMORPG seems very important for the game company. Furthermore, because game users belong to some parties to play and hunt together, setting an adequate level of gaming difficulty is very important for the game company to attract voluntary participation of gaming parties from users.

- Discussion

There are few limitations of the study to make mention of. the classic academic view on the Flow Theory has embrace the difficulty conditions regarding to user's experience on its circumstances. Therefore, some part of this research is open to dispute by other scholars. However, because the concept of difficulty conditions may be substituted by the party counts that point out how many times the users get together to play and hunt

in the MMORPG, it seems there is no problems for readers to interpret and understand. On the contrary, it can fairly be said that quite a lot of parts of this study can be read in diverse ways, and it actually expand the view of traditional Flow Theory.

For its further research, a prediction analysis methodology could be a good attempt with the logistic analysis of the research. Because the study is focused on the effects comparing groups with hierarchical analysis, the research cannot approach a new topic like the prediction analysis. Since this research wants to avoid a study for the sake of study, it is hoping that this suggestion walks you through a great guideline.

References

Csikszentmihalyi, M., & Nakamura, J. (1979). The concept of flow. *Play and learning*, 257-274.

Charlton, J. P., & Danforth, I. D. W. (2007). Distinguishing addiction and high engagement in the context of online game playing. *Computers in Human Behavior*, *23*, 1531–1548.

Choi, D., & Kim, J. (2004). Why people continue to play online games: In search of critical design factors to increase customer loyalty to online contents. *CyberPsychology & Behavior*, *7*, 11–24.

Chou, C., & Hsiao, M. C. (2000). Internet addiction, usage, gratification, and pleasure experience: The Taiwan college students' case. *Computers & Education*, *35*, 65–80.

Chou, T. J., & Ting, C. C. (2003). The role of flow experience in cyber-game addiction. *CyberPsy- chology & Behavior*, *6*, 663–675.

Greene, William H. (2008). Econometric Analysis (Sixth ed.). Upper Saddle River: Pearson Prentice-Hall. pp. 314–353

Griffiths, M. D., Davies, M. N. O., & Chappell, D. (2004). Demographic factors and playing variables in online computer gaming. *CyberPsychology & Behavior*, *7*, 479–487.

Gujarati, Damodar N.; Porter, Dawn C. (2009). Basic Econometrics (5th.). New York: McGraw-Hill Irwin. pp. 711–736

Holt, N. A., & Kleiber, D. A. (2009). The Sirens' song of multiplayer games. *Children, Youth and Environments*, *19*, 223–244.

Howell, D. C. (2012). *Statistical methods for psychology*. Cengage Learning.

Hsu, C. L., & Lu, H. P. (2004). Why do people play on-line games? An e extended TAM with social influences and flow experience. *Information & Management*, *41*, 853–868.

Morahan-Martin, J., & Schumacher, P. (2000). Incidence and correlates of pathological Internet use among college students. *Computers in Human Behavior*, *16*, 13–29.

Nakamura, J., & Csikszentmihalyi, M. (2009). Flow theory and research. *Handbook of positive psychology*, 195-206.

Wen, M. H., Hsu, S. H., & Wu, M. C. (2009). Exploring user experiences as predictors of MMORPG addiction. *Computers & Education*, *53*, 990–999.

Wooldridge, Jeffrey M. (2013). Introductory Econometrics: A Modern Approach (Fifth international ed.). Mason, OH: South-Western. pp. 490–528.

Yoo, B., Kim, K., & Kauffman, R. J. (2013). Valuation of participation in social gaming. *International Journal of Electronic Commerce*, *18*(2), 11-50.

국문초록

온라인 게임 안에서 각 개인의 사용자는 몰입 정도에 따라 미시적인 행태변화 (예를들면, 구매행태, 플레이행태, 네트워크 교류 행태 등) 가 나타나게 되는데 이는 게임의 난이도 수준을 조정할 때 매우 중요한 지표가 될 뿐만 아니라, 게임 회사 입장에서 이는 매출과 직결되는 사용자의 구매행위에 영향을 미치게 되므로 매우 주요한 연구 사항이라 할 수 있다.

따라서, 본 연구에서는 사용자 개개인의 데이터를 분석함으로써 미시적인 행태변화에 따른 긍정적 또는 부정적 영향도를 살펴보고자 한다. 첫째, 심리학에서 주로 사용되는 몰입이론 (Flow Theory)을 차용하여 전체 표본의 행태를 파악하고자 회귀분석을 시행하였다. 둘째, 성인과 청소년 그룹간 다각적인 비교를 위해 그룹간 다중회귀 분석을 시행하였다. 셋째, 매출과 직결되는 사용자의 구매행태 분석을 위해 구매에 영향을 미치는 요인을 파악하여 로짓 분석을 시행하였다. 끝으로, 추후 연구 방안을 제시함으로써 선행연구의 한계점을 짚고 향후 연구의 방향을 논하였다.